**Science Cartography:**
**Communicating, Navigating, Managing, and Utilizing our Collective Scholarly Knowledge Across Disciplinary, Cultural, and Geospatial Boundaries**

I believe it would be highly beneficial if 13 Federal agencies would collaborate on federating the existing publication, patent, funding datasets with the goal to make this vast amount of data/knowledge and expertise easy to understand, navigate, manage, and utilize. The resulting dataset can also be used to analyze, model, and (visually) communicate the structure and dynamics of science and technology in support of science policy and economic decision making. A more detailed argumentation and discussion of benefits for different user groups follows below.

The number of currently active researchers exceeds the number of researchers ever alive. Researchers publish or perish. Some areas of science produce more than 40,000 papers a month. We are expected to know more works than one can possibly read in a lifetime. We receive many more emails per day than can be processed in 24 hours. We are supposed to be intimately familiar with datasets, tools, and techniques that are continuously changing and increasing in number and complexity. All this while being reachable twenty four hours, seven days a week.

Not only library buildings and storage facilities, but also databases are filling up quicker than they can be built. In addition, there are scientific datasets, algorithms, and tools that need to be mastered in order to advance science. The center figure depicts just how much information exists. No single man or machine can process and make sense of this enormous stream of data, information, knowledge, and expertise.

All this leads to a quickly increasing specialization of researchers, practitioners, and other knowledge workers; a disconcerting fragmentation of science; a world of missed opportunities for collaboration; and a nightmarish feeling that we are doomed to 'reinvent the wheel' forever.

Over the last several hundred years, our collective knowledge has been preserved and communicated via scholarly works such as papers or books. Works might report a novel algorithm or approach, report experimental results, or review one area of research among others. Some report several small results others one large result like the Human Genome. Some confirm results while others disprove results. Some are filled with formulas while others feature mostly artistic imagery. Different areas of science have very different publishing formats and quality standards. The description of one and the same result, e.g., a novel algorithm, will look very different if published in a computer science, biology, or physics journal.

Unfortunately, our tools to access, manage, and utilize our collective knowledge are rather primitive. In fact, our main means of accessing everything we collectively know are search engines. Search engine providers try to make us believe that we can live in 'flatland'. That is, no directory structures are necessary and data objects can have any name. Superior retrieval software will find whatever we need. However, the usage of search engines resembles charging a needle with a search query and sticking it into a haystack of unknown size and consistency. Upon pulling the needle out, one checks the linearly sorted items that got stuck on it. This seems to work well for fact-finding. However, it keeps us always at the bottom of confirmed and unconfirmed records in the haystack of our collective knowledge. Of course, we can explore local neighborhoods of retrieved records via Web links or citation links. However, there is no 'up' button that provides us with a more global view of what we collectively know and how everything is interlinked. Search engines are a bad choice when it comes to identifying patterns, trends, outliers, or the context in which a piece of knowledge was created or can be used. Without context, intelligent data selection, prioritization, and quality judgments become extremely difficult.

Consequently, even the smartest people on this planet can neither keep up with the amount nor the accelerating speed of knowledge production. This becomes a major concern if scientific results are needed to enable all human beings to live a healthy, productive, and fulfilling life. We simply need better tools to keep track, access, manage, and utilize our collective scholarly knowledge and expertise.

Maps of science that guide our scholarly endeavors promise to make a difference. They aim to serve today's explorers navigating the world of science. These maps are generated through scientific analysis of large-scale scholarly datasets in an effort to connect and make sense of the bits and pieces of knowledge they contain. They can be used to objectively identify major research areas, experts, institutions, collections, grants, papers, journals, and ideas in a domain of interest. Local maps provide overviews of a specific area: its homogeneity, import-export factors, and relative speed. They allow one to track the emergence, evolution, and disappearance of topics and help to identify the most promising areas of research. Global maps show the overall structure and evolution of our collective scholarly knowledge.

Science maps have been designed for diverse users and information needs as discussed subsequently.

### *Science Maps as Visual Interfaces to Scholarly Knowledge*

The number and variety of available databases in existence today is overwhelming. Databases differ considerably in their temporal, geospatial, and topical coverage. Database quality reaches from 'downloaded from the Web' to 'manually curated by experts'. Visual interfaces to digital libraries provide an overview of a library's or publisher's holdings as well as an index into its records. They apply powerful data analysis and information visualization techniques to generate visualizations of large document sets. The visualizations are intended to help humans mentally organize, electronically access, and manage large, complex information spaces and can be seen as a value-adding service to digital libraries.

### *Mapping Intellectual Landscapes for Economic Decision Making*

A deep understanding of technology, governmental decisions, and societal forces is required to make informed economic decisions that ensure survival in highly competitive markets. Discontinuities caused by disruptive technologies have to be determined and relevant innovations need to be detected, deeply understood, and exploited. Companies need to look beyond technical feasibility to identify the value of new technologies, to predict diffusion and adoption patterns, and to discover new market opportunities as well as threats.

The absorptive capacity of a company, i.e., its ability to attract the best brains and 'to play with the best' has a major impact on its survival. The importance of social networking tools and network visualizations increases with the demand to understand the "big picture" in a rapidly changing global environment.

Competitive technological intelligence analysis, technology foresight studies, and technology road mapping are used to master these tasks. Easy access to major results, data, tools, expertise is a key to success. But exactly what is the competition doing, who makes what deals, and what intellectual property rights are claimed by whom?

Last but not least, companies need to communicate their image and goals to a diverse set of stakeholders -- to promote their products, to hire and cultivate the best experts, and to attract venture capital.

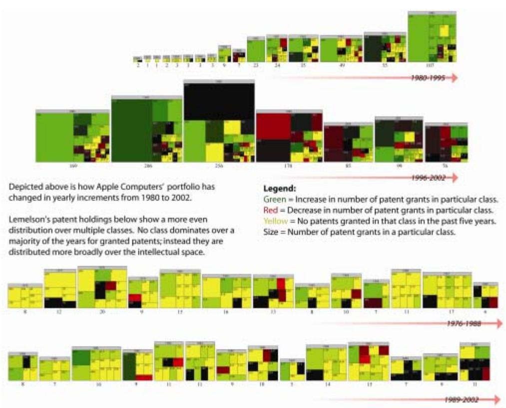


*Figure 1 Claiming Intellectual Property Rights via Patents*
*The evolving patent portfolios of Apple Computer (1980-2002) and Jerome Lemelson (1976-2002) are shown here. The number of patents granted per year matches the size of the square. Each square is further subdivided into patent classes which are color coded in green if the number of patents increased, in red if it decreased, and yellow if no patent was granted in this class in the last five years. While Apple Computer claims more and more space in the same classes, Lemelson's patent holdings are distributed more broadly over the intellectual space.*

### *Science of Science Policy (Maps) for Government Agencies*

Increasing demands for accountability require decision makers to assess outcomes and impacts of science and technology policy. There is an urgent need to evaluate the impact of funding/research on scientific progress: to monitor (long-term) money flow and research developments, to evaluate funding strategies for different programs, and to determine project durations and funding patterns.

Professional science managers are interested to identify areas for future development, to stimulate new research areas, and to increase the flow of ideas into products. Hence, they need to identify emerging research areas; understand how scientific areas are linked to each other; examine what areas are multi-disciplinary, measure collaborations and knowledge flows at the personal, institutional, national, and global level; identify and compare core competencies of economically competing institutions and countries; identify and fund central and not peripheral research centers, etc.

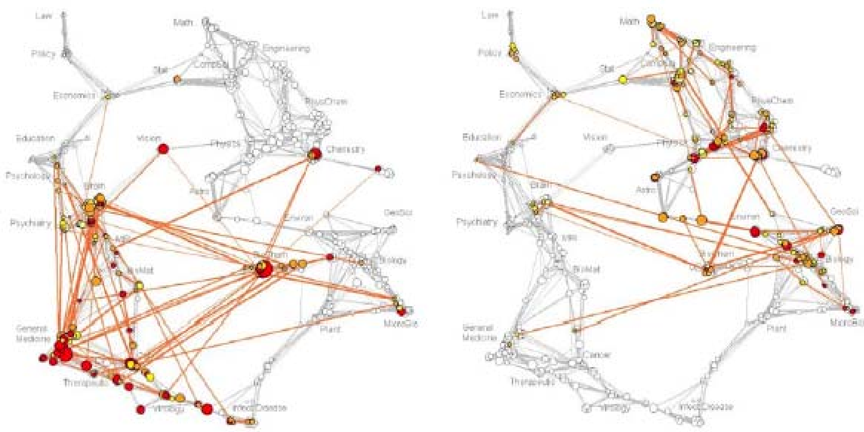

*Figure 2: Funding Profiles of NIH (left) and NSF (right)*
*Using a base map of science, the core competency of institutes, agencies, or countries can be mapped and visually compared. Shown here are funding profiles of the National Institutes of Health (NIH) and the National Science Foundation (NSF). As the base map represents papers, funding was linked by matching the first author of a paper and the principal investigator using last name and institution information. A time lag of three years between funding of the grant and publication of the paper was assumed. While NIH mostly funds biomedical research, NSF focuses on math, physics, engineering, computer science, environmental and geo sciences, and education. Overlaps exist in chemistry, neuroscience, and brain research.*

### *Professional Knowledge Management Tools for Scholars*

Most researchers wear multiple hats: They are researchers, authors, editors, reviewers, teachers, mentors, and often also science administrators.

As researchers and authors, they need to strategically exploit their expertise and resources to achieve a maximum increase of their reputation. Expertise refers to the knowledge they already have or can obtain in a given time frame but also expertise that can be acquired via collaborations. Resources refer to datasets, software, and tools but also to people supervised or paid.

They need to keep up with novel research results; examine potential collaborators, competitors, related projects; weave a strong network of collaborations; ensure access to high quality resources; and monitor funding programs and their success rates. Last but not least, they need to review and incorporate findings and produce and diffuse superior research results in exchange of citation counts, download activity, press, etc.

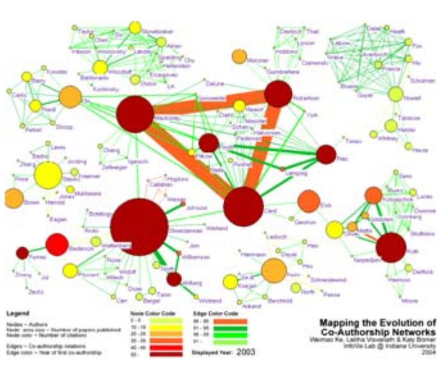


*Figure 3: Mapping the Evolution of Co-Authorship Networks*
*This is the last frame of an animation sequence that shows the evolution of authors (nodes) and their co-authorship relations (links) in the domain of information visualization. Node area size coded reflects the number of papers an author published, its color denoted the number of citations these papers received.*
*Link width equals the number of co-authorships and link color denotes the year of the first collaboration. Large, dark nodes are preferable. Large, light nodes indicate many papers that have not (yet) been cited. Shneiderman -- working in a student dominated academic setting – has a very different co-author environment than Card, Mackinley, and Robertson which are at Xerox Parc for most of the time captured here.*

As editors and reviewers, researchers act as gatekeepers of science. They need detailed expertise of their domain as well as related domains of research to ensure that only the most valuable and unique works become manifested in the eternal pile mountain of scholarly works.

As teachers and mentors, they provide students with a deep understanding of the structure and evolution but also the peculiarities of a domain of research and practice. They might give an overview of the teaching material to be covered first; then highlight major people, important papers, and key events; and provide a red thread or pathway and help students discover, understand, and interrelate details.

As science administrators, they are responsible for decisions regarding hiring and retention; promotion and tenure; internal funding allocations; budget allocation; outreach. Here a global overview of major entities and processes and their temporal, spatial, and topical dynamics is important.

***Science Maps for Kids***

In school, children learn about many different sciences. However, they never get to see science 'from above'. Hence, they have a very limited understanding regarding the sizes of different sciences or their complex interrelations.

How will they be able to answer questions such as: What intellectual travels did major inventors undertake to succeed? Why is mathematics necessary to succeed in almost all sciences? How do the different sciences build on each others work? Or how would they find their place in science and where would it be?

Imagine a map of our collective scholarly knowledge would hang next to the map of the world in each class room. Imagine students could not only travel our planet online but also explore the web of knowledge. How would this change the way we learn, understand, and create?

See also http://scimaps.org/kids/Web_topic_big.jpg - different base map as funding overlay

*Figure 4: Hands-on Science Maps for Kids invite children to see, explore, and understand science from above. This map shows our world and the places where science is practiced or researched. A complementary map shows major areas of science and their complex interrelationships. Children and adults alike are invited to help solve the puzzle by placing major scientists, inventors, and inventions at their proper places. Look for the many hints hidden in the drawings to find the perfect place for each puzzle piece. What other inventors and inventions do you know? Where would your favorite science teachers and science experiments go? What area of science do you want to explore next?*

**References**

Chen, C. (2002) *Mapping Scientific Frontiers.* Springer-Verlag, London.

Börner, K., C. Chen, and Boyack, K. W. (2003) Visualizing Knowledge Domains, in *Annual Review of Information Science & Technology*, B. Cronin, Editor. Information Today, Inc./American Society for Information Science and Technology: Medford, NJ, pp. 179-255.

Shiffrin, R. M. and Börner, K. (eds.) (2004) *Mapping Knowledge Domains*. PNAS. Vol. 101 (Suppl. 1). PNAS.

Boyack, K.W., R. Klavans, and Börner, K. (2005) Mapping the Backbone of Science. *Scientometrics*, **64** (3), 351-374.

Börner, Katy, Dall'Asta, Luca, Ke, Weimao & Vespignani, Alessandro. (2005). Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-Authorship Teams. *Complexity*. Vol. 10*(4)*, 58-67.

Börner, Katy. (2009). Atlas of Science: Guiding the Navigation and Management of Scholarly Knowledge. ESRI Press.

Mapping Science Exhibit online at http://scimaps.org and on display at the National Science Foundation, 4201 Wilson Boulevard, Arlington, VA.